



Published in final edited form as:

*Am J Gastroenterol.* 2022 October 01; 117(10): 1573–1582. doi:10.14309/ajg.0000000000001871.

## OPTIMAL WIRELESS REFLUX MONITORING METRICS TO PREDICT DISCONTINUATION OF PROTON PUMP INHIBITOR THERAPY

Rena Yadlapati, MD MS<sup>1</sup>, C. Prakash Gyawali, MD<sup>2</sup>, Melina Masihi, PhD<sup>3</sup>, Dustin A Carlson, MD MS<sup>3</sup>, Peter J. Kahrilas, MD<sup>3</sup>, Billy Darren Nix, BS<sup>2</sup>, Anand Jain, MD<sup>4</sup>, Joseph R. Triggs, MD PhD<sup>5</sup>, Michael F. Vaezi, MD PhD<sup>6</sup>, Leila Kia, MD<sup>3</sup>, Alexander Kaizer, PhD<sup>7</sup>, John E. Pandolfino, MD MS<sup>3</sup>

<sup>1</sup>Division of Gastroenterology, University of California, San Diego School of Medicine, La Jolla, CA, USA

<sup>2</sup>Division of Gastroenterology, Washington University School of Medicine, St. Louis, MO, USA

<sup>3</sup>Division of Gastroenterology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>4</sup>Division of Digestive Diseases, Emory University School of Medicine, Atlanta, GA, USA

<sup>5</sup>Division of Gastroenterology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>6</sup>Division of Gastroenterology, Hepatology and Nutrition, Vanderbilt University Medical Center, Nashville, TN, USA.

<sup>7</sup>University of Colorado, Anschutz Medical Campus, Colorado School of Public Health, Department of Biostatistics and Informatics, Aurora, CO, USA

### Abstract

**Background:** Ambulatory reflux monitoring performed off proton pump inhibitor (PPI) is the gold standard diagnostic test for non-erosive gastroesophageal reflux disease (GERD). However, the diagnostic metrics and optimal duration of monitoring are not well defined. This study evaluated the performance of multiple metrics across distinct durations of wireless reflux monitoring off PPI against the ability to discontinue PPI therapy in patients with suboptimal PPI response.

**Methods:** This single-arm clinical trial performed over four years at two centers enrolled adults with troublesome GERD symptoms and inadequate response to 8 weeks of PPI. Participants underwent 96-hr wireless pH monitoring off PPI. Primary outcome was whether the subject successfully discontinued PPI or resumed PPI within 3 weeks.

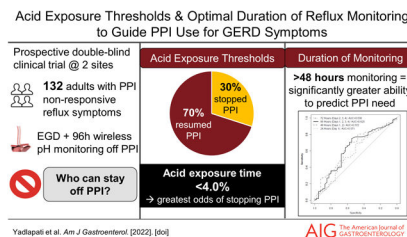
**Corresponding Author:** Rena Yadlapati MD MS, 9500 Gilman Drive MC0956, La Jolla, CA 92093, ryadlapati@health.ucsd.edu.

**Specific Author Contributions:** Study design/conception: RY, JEP; Acquisition of data: MM, DAC, CPG, PJK, BDN, JT, AJ, MFV, LK, JEP; Analysis of data: RY, AK, JEP; Interpretation of data: RY, AK, JEP; Drafting of manuscript: RY, MM, AK, JEP; Critical revision of manuscript: RY, MM, DAC, CPG, PJK, BDN, JT, AJ, MFV, LK, AK, JEP; Final approval of manuscript to be published: RY, MM, DAC, CPG, PJK, BDN, JT, AJ, MFV, LK, AK, JEP.

**Results:** Of 132 participants, 30% discontinued PPI. Among multiple metrics assessed, total acid exposure time (AET) of 4.0% performed best in predicting PPI discontinuation (OR 2.9 (95% CI 1.4, 6.4);  $p=0.006$ ), with other thresholds of AET and DeMeester score performing comparably. AET was significantly higher on day 1 of monitoring compared to other days, and prognostic performance significantly declined when only assessing the first 48-hrs of monitoring (Area under curve (AUC) for 96 hours 0.63 vs AUC for 48 hours 0.57;  $p=0.01$ ).

**Conclusion:** This clinical trial highlights the AET threshold of 4.0% as a high performing prognostic marker of PPI discontinuation. 96-hrs of monitoring performed better than 48-hr, in predicting ability to discontinue PPI. These data can inform current diagnostic approaches for patients with GERD symptoms that are unresponsive to PPI therapy

## Graphical Abstract



## Keywords

Functional heartburn; non-cardiac chest pain; erosive esophagitis; Barrett's esophagus; esophageal motility

## INTRODUCTION

Nearly half of patients experiencing esophageal symptoms suspicious for gastro-esophageal reflux disease (GERD) remain symptomatic despite proton pump inhibitor (PPI) therapy.<sup>3-6</sup> Current evidence-based recommendations support esophageal testing for objective GERD in symptomatic PPI non-responder patients, beginning with an upper gastrointestinal endoscopy and followed by ambulatory reflux monitoring off acid suppression in the absence of conclusive endoscopic evidence of GERD.<sup>1, 2</sup> Ambulatory reflux monitoring is the current gold standard for diagnosis of non-erosive GERD, which measures distal esophageal acid exposure time (AET) and determines correlation with patient reported symptoms.<sup>3</sup> Wireless reflux monitoring has several advantages in this context, including placement during the same endoscopy session when endoscopy is 'negative', prolonged acid monitoring capacity for up to 96 hours, and increased diagnostic yield for GERD.<sup>4-9</sup> As many as 30 to 75% of patients undergoing prolonged wireless reflux monitoring off acid suppression will have physiologic acid exposure indicating a negative study, suggesting a non-GERD mechanism of symptoms,<sup>10, 11</sup> and leading to successful discontinuation of PPI therapy without negative consequences.<sup>12</sup> In the current clinical environment of widespread and often inappropriate PPI usage, this ability to identify individuals in whom PPI therapy can be discontinued is a distinct advantage. Thus, ambulatory reflux monitoring, particularly prolonged wireless reflux monitoring, is a valuable diagnostic and prognostic tool in the clinical evaluation of persisting esophageal symptoms suspicious for GERD.

Despite its obvious clinical value, wireless reflux monitoring is often performed and interpreted variably across centers. There are several knowledge and evidence gaps within the context of wireless reflux monitoring. First, metrics and thresholds used to guide clinical impression, including AET and reflux-symptom association are not standardized, and are used variably as composite vs. daily metrics, with variable upper limits of normal. Second, the optimal duration of wireless reflux monitoring remains incompletely studied, and the value of extending monitoring beyond 48 hours has not been conclusively determined. Finally, best practice updates and guidelines refrain from endorsing firm thresholds, since supporting data constitute low-level evidence based on observational studies and expert consensus. Consequently, well-designed clinical trials are critically needed to address these knowledge gaps.

The current double-blind prospective clinical trial assessed the performance of wireless reflux monitoring performed off acid suppression in symptomatic patients using a multitude of metrics across distinct durations of monitoring, using the ability of the patient to discontinue PPI therapy as an outcome variable.

## METHODS

### Study Design & Aims

This double-blind single-arm prospective clinical trial was conducted over four years (May 2017 to September 2021) at two tertiary-care centers (Lead site: Northwestern University, Chicago, IL; Second site: Washington University, St. Louis, MO). Inclusion criteria consisted of adults with gastro-esophageal reflux symptoms and inadequate response to PPI therapy presenting to outpatient clinics at the two centers for clinical evaluation, with no prior conclusive endoscopic reflux esophagitis, and without prior foregut surgery. In this study, participants were advised to cease PPI use for 3 weeks, in order to observe the rate of relapse to PPI use. The analysis presented in this paper focuses on performance characteristics of prolonged wireless reflux monitoring compared to the reference standard of ability to discontinue PPI therapy. The study was approved by the Institutional Review Board at each site and was registered with [clinicaltrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT03202537) NCT03202537.

### Study Population

Adult patients with troublesome esophageal symptoms (at least two episodes of heartburn, regurgitation and/or non-cardiac chest pain per week) who remained symptomatic despite a compliant trial of at least single-dose PPI therapy for a minimum of 8 weeks were eligible for enrollment. Exclusionary criteria included the presence of confirmatory erosive GERD including active severe erosive esophagitis (Los Angeles Grade C or D) on endoscopy or long-segment Barrett's esophagus (> 3cm in length) as per Lyon Consensus criteria. Exclusionary criteria also included prior foregut surgery, signs or symptoms of active heart disease, pregnancy, manometric evidence of a major motility disorder (according to Chicago Classification<sup>13</sup>), or >15 eosinophils per high power field on esophageal biopsies obtained when endoscopic findings raised suspicions of eosinophilic esophagitis. Patients experiencing isolated extra-esophageal symptoms in the absence of esophageal symptoms were excluded. Patients with insufficient pH monitoring time captured (at least 14 hours/day

for 3 days) were also excluded. All participants provided written informed consent prior to study enrollment, esophageal physiologic testing, and symptom evaluation. There is overlap with the cohort described in a previous report.<sup>12</sup>

### Study Protocol & Intervention

The study intervention consisted of instruction to participants to cease PPI use, in order to determine ability to remain off PPI over 3 weeks. During this three week study period participants underwent 96-hour wireless reflux monitoring after being off PPI for at least one week. Participants were instructed to continue to refrain from resuming PPI therapy for an additional two weeks following reflux monitoring. During these two weeks participants could use over the counter antacids (e.g., Tums, Rolaids) up to 5 times per day if needed for symptom relief. Indication to resume PPI was defined as participant report of high symptom burden with a desire to resume PPI, and/or in excess of the maximal over the counter antacid utilization. The research coordinator contacted participants every week during the study to assess PPI resumption; additionally, participants were advised to e-mail or call the research coordinator if they met these criteria. Participants, as well as study investigators, were blinded to results of reflux testing during the intervention.

**Wireless pH Monitoring**—During sedated upper gastrointestinal endoscopy, the wireless pH probe delivery catheter (Bravo; Medtronic, Minneapolis, MN) was introduced transorally and the pH capsule was positioned 6 cm proximal to the endoscopically identified squamocolumnar junction, corresponding to 5 cm above the proximal border of the lower esophageal sphincter. Once the catheter was in the appropriate position, the external portable vacuum pump was switched on to apply suction to the well of the capsule and suction in adjacent esophageal mucosa. After 30 seconds, the plastic safety guard was removed, and the activation button was depressed. Participants were instructed to continue usual daily activities and meals, remain off PPI, and log symptoms/meals in a written and electronic diary, while remaining within 3 feet of the pager-sized receiver at all times. Participants returned the wireless pH study receiver 96 hours later, at which point data were downloaded and analyzed using a proprietary commercial interpretation platform (Reflux Reader, Medtronic, Minneapolis, MN).

**Patient Reported Symptoms**—At enrollment, all participants completed the GERDQ instrument<sup>14, 15</sup> as well as the Reflux Symptom Questionnaire electronic Diary during the daytime as well as nighttime, at four time points during the study: on PPI at time of enrollment and off PPI at weeks one, two and three.<sup>16</sup> The study coordinator contacted participants weekly for four weeks to monitor symptoms, collect questionnaire scores, and determine whether PPI therapy was resumed.

### Data Source & Measurement

Data for all participants were electronically collected in a uniform de-identified dataset through Research Electronic Data Capture hosted at the lead study site with multi-site access for the secondary participating center. Data collected for participants included demographics, endoscopic findings (presence/degree of erosive esophagitis, hiatal hernia size), eosinophil count on esophageal biopsy, questionnaire scores, and PPI use. Reflux

monitoring data analyzed by a blinded external investigator using manufacturer software (Reflux Reader; Medtronic, Minneapolis, MN) included monitoring time, total, upright, supine, and daily acid exposure time (AET; percent time esophageal acid exposure is below pH of 4.0), DeMeester score, number of reflux events, longest reflux event, symptoms reported, symptom index (proportion of symptoms associated with a reflux episode; optimal threshold > 50%) and symptom association probability (a statistical calculation expressing the probability that symptom events and reflux episodes are associated; 95% considered positive).<sup>10</sup> Acid exposure was further stratified into dominant and discordant patterns (physiologic dominant pattern, pathologic dominant pattern, borderline dominant pattern or discordant pattern), and trajectory of acid exposure was determined (high, mid or low-exposure trajectory), as described in prior reports.<sup>14, 15</sup>

## Data Analysis

Continuous data are described by mean (standard deviation) and categorical data as frequency (percent), unless otherwise indicated. The primary analysis aimed to assess performance of data from prolonged wireless reflux monitoring against the outcome of ability vs. inability to discontinue PPI using receiver operating characteristics (ROC), with the area under the curve and its 95% confidence intervals calculated via DeLong's method. AET was examined across thresholds from 0 to 14 in increments of 0.1 to identify the value that achieved 90% sensitivity or specificity in predicting PPI discontinuation. Odds ratios were estimated with logistic regression. Linear regression was used to compare the average daily AET across the four days of data collection using generalized estimating equations assuming a first-order autoregressive working correlation structure to account for the daily measurements. All figures and analyses were conducted using R v4.1.0 (Vienna, Austria), with ROC curves generated using the pROC package, and sensitivity and specificity calculated from the epiR package.<sup>16, 17</sup>

## RESULTS

### Baseline Characteristics

A total of 132 participants are included in the analysis (Figure 1): 53 (40%) male, mean age 47.3 years (SD 14.6), and mean body mass index 27.1 kg/m<sup>2</sup> (SD 5.6). Mean wireless reflux monitoring time was 3.4 days (SD 0.4). Total AET was 4.0% or less in 49 (37%) participants and greater than 4.0% in 83 (63%), and greater than 6.0% in 60 (45%) of participants. Of the 132 participants, 40 (30%) were able to discontinue PPI use, while 92 (70%) resumed PPI.

### Performance Characteristics of Acid Exposure Thresholds from 96-Hour Wireless Reflux Monitoring (Tables 1 & 2)

**Overall Acid Exposure Time Thresholds:** An AET threshold of 3.95% demonstrated optimal overall performance for PPI discontinuation [AUC 0.63 (95% CI 0.52, 0.73)] with 75% sensitivity and 55% specificity. Each total AET threshold of 4.0%, 5.0% and 6.0% (based on the Lyon Consensus)<sup>3</sup> significantly predicted PPI discontinuation, with total AET less than 4.0% having the greatest odds of predicting PPI discontinuation [OR 2.9 (95% CI 1.4, 6.4); p=0.006]. Specifically, 45% (22/49) with total AET < 4.0% discontinued PPI compared to 22% (18/83) with total AET > 4.0%. The lowest AET with at least 90%

sensitivity was an overall AET of 1.2% (91.3% sensitivity, 12.5% specificity) and the highest AET with at least 90% specificity was an overall AET of 10.3% (90.0% specificity, 14.1% sensitivity) in predicting PPI discontinuation.

**Number of days with elevated AET:** Performance characteristics for the number of days with AET >4.0%, 5.0%, or 6.0%<sup>3</sup> were assessed. The number of days with AET greater than 4.0% had the greatest performance for predicting PPI discontinuation [AUC 0.65 (95% CI 0.55, 0.75); 70% sensitivity, 55% specificity]. Specifically, 54% (13/24) with 0 days of an AET > 4.0% discontinued PPI compared to 19% (7/37) with 4 days of AET >4.0% [OR 5.0 (95%CI 1.7, 16.7); p=0.006]. The optimal number of days with a specific AET threshold was as follows for predicting outcome of PPI discontinuation: 1.5 days for AET > 4.0%, 0.5 days for AET > 5.0%, and 0.5 days for AET >6.0%.

**DeMeester Score:** Overall, the continuous DeMeester score, a composite of AET, number of reflux events, and longest reflux event, performed similarly to total AET [AUC 0.62 (0.52, 0.73)] in predicting PPI discontinuation. Patients with a DeMeester score of 14.2 or less had a 2.6 increased odds (95% CI 1.1, 6.4; p=0.032) of PPI discontinuation compared to patients with a DeMeester score >14.2. Patients with a DeMeester score of 50 or less had a 4.6 fold increased odds (95% CI 1.2, 30.3; p=0.048) of PPI discontinuation compared to patients with a DeMeester score greater than 50.

**Acid Exposure Patterns:** In accordance with prior published studies, AET data were categorized according to a dominant pattern [physiologic dominant (AET < 4.0% for at least 2 days), pathologic dominant (AET > 6.0% for at least 2 days), or borderline/discordant if not meeting criteria for physiologic or pathologic dominant]. PPI discontinuation was significantly greater with a physiologic dominant pattern compared to pathologic dominant pattern (OR 2.3 (95% CI 1.1, 5.2); p=0.038).

Trajectory analysis identified three trajectories of acid exposure (low, mid and high) similar to prior published data.<sup>15</sup> PPI discontinuation was significantly higher among low trajectory acid exposure compared to mid trajectory (OR 2.3 (95% CI 1.0, 5.4); p=0.048). Only 9 subjects met criteria for high trajectory, 78% of whom resumed PPI, though statistical comparisons were limited given the small sample size.

Performance of symptom index and symptom association probability are presented in Table 1.

### Performance of AET Across Varying Durations of Wireless Reflux Monitoring

Performance of various durations of reflux monitoring including 24 hours alone (day 1), 48 hours (day 1 and 2), exclusion of day 1 (day 2, 3, and 4) were assessed and compared to overall 96 hour monitoring data.

**Acid Exposure Across Various Days of Monitoring:** Mean AET day by day was as follows: day 1: 7.2% (95% CI: 6.1, 8.3), day 2: 5.7% (95% CI: 4.9, 6.5), day 3: 5.6% (95% CI: 4.8, 6.4), and day 4: 5.3% (95% CI: 4.6, 6.0). AET on day 1 was significantly greater across all participants relative to day 2 (p=0.001), day 3 (p=0.003), and day 4 (p<0.001).



Specifically, on further analysis of 62 patients with an AET > 6.0% on day 1, AET was discordant (less than 6.0%) in 25 patients (40%) on day 2, 29 patients (47%) on day 3, and 31 patients (50%) on day 4. Across all participants AET did not differ between days 2, 3 or 4 ( $p>0.05$  for all comparisons).

### **Performance Characteristics of AET Across Varying Durations of**

**Monitoring:** Figure 2 depicts the AUC for performance of AET in predicting PPI discontinuation by duration of monitoring. AUC was similar between 24 hour (day 1) monitoring and 48 hour (day 1 and 2) monitoring ( $p=0.49$ ). AUC was significantly greater for 96 hours compared to 24 hour of monitoring ( $p=0.03$ ) or 48 hour of monitoring ( $p=0.01$ ). Figure 3 compares the risk estimates of PPI discontinuation when data were assessed across 96 hours, 48 hours, or 96 hours with day 1 excluded.

**48-Hour Monitoring:** Since 48 hour wireless pH monitoring is commonly utilized we include a complete analysis of performance of data from 48 hour monitoring in Table 3. Notably, assessment of the number of days over a 48-hour monitoring period with an elevated AET did not identify significant differences between proportions of patients able to discontinue PPI. For instance, there was no significant difference in PPI withdrawal between patients with 0 days of AET < 4.0% (15/37, 41%) or patients with 2 days of AET > 4.0% (15/62, 24%;  $p=0.090$ ). If we consider data from only day 2 of monitoring, AET of 4.0% performed with 60% sensitivity (95% CI 49%, 70%) and 52% specificity (36%, 68%) in predicting PPI discontinuation. As depicted in Figure 2, AUC of day 2 AET (0.566 (95% CI 0.46, 0.67) was similar to that of AET from day 1 of monitoring and 48 hours (day 1 and 2) monitoring

**96-Hour Monitoring Excluding Day 1 (Including only Days 2, 3, and 4):** Given concerns around reliability of day 1 AET measurements we include a complete analysis of performance data from 96 hour monitoring with day 1 excluded in Table 4. Notably, overall performance improved when excluding day 1 of monitoring from the analysis (AUC 0.64 [95% CI 0.53, 0.74]). The AUC of 96 hours monitoring excluding day 1 was similar to the AUC of 96 hours monitoring total ( $p=0.43$ ).

## **DISCUSSION**

Despite common utilization of wireless reflux monitoring in the evaluation of GERD symptoms, the performance of diagnostic thresholds and duration of wireless reflux monitoring had hitherto not been studied against a clinically relevant outcome.<sup>18</sup> Consequently, uncertainty surrounding thresholds and interpretation of wireless pH data has led to variability in diagnostic impressions and management decisions. In this first-of-its-kind prospective clinical trial, we assessed performance characteristics of clinically relevant metrics from wireless reflux monitoring against the ability to discontinue PPI therapy in 132 patients with esophageal symptoms suspicious for GERD and incomplete PPI response. We demonstrate the high performing prognostic value of the AET threshold of <4.0% in predicting PPI discontinuation, indicating low likelihood of reflux mediated symptoms. At the opposite end of the spectrum, AET >10.0% and/or DeMeester score >50.0 had similar high performance in predicting need for chronic anti-reflux therapy. Further, we

describe the incremental value of 96-hour recordings over 48-hour recordings in predicting PPI discontinuation, even when the first day of the recording (which may be fraught with procedure/sedation related variation in acid exposure) is excluded from analysis. These findings demonstrate the clinical utility of 96-hour wireless reflux monitoring performed off acid suppression and establish AET thresholds for making a diagnosis in patients with esophageal symptoms suspicious for GERD, as well as predicting need for discontinuation of acid suppression in non-responders.

For more than four decades, AET has been extensively studied and widely accepted as a reproducible measure extracted from ambulatory reflux monitoring.<sup>19</sup> However, with the advent of prolonged wireless reflux monitoring, observations of variability in day-to-day acid exposure raised uncertainty regarding optimal interpretation of reflux metrics beyond 24 hours.<sup>18, 20</sup> Consequently, investigators have proposed sophisticated assessments such as the “dominant pattern”<sup>14</sup> or “trajectory analysis”<sup>15</sup>, or have rejected the value of prolonging acid measurements altogether.<sup>4</sup> In addition to high quality evidence that esophageal acid exposure burden is a clinically relevant physiologic marker of gastro-esophageal reflux burden, this first-of-its-kind double-blinded clinical trial<sup>12</sup> demonstrates the comparable, and in some cases better performance of a simple assessment of daily acid exposure from multiple days of recording compared to other composite or complex assessments. This study further validates the Lyon Consensus recommendation of AET <4.0% off acid suppression<sup>3</sup> as physiologic acid exposure warranting discontinuation of acid suppression to curb unwarranted and unnecessary PPI use. While the precise AET threshold for a definitive diagnosis of pathologic acid exposure remains unclear, these data suggest that AET > 5.0% or 6.0% increase diagnostic specificity, and greater acid exposure times (>10.0% in this study) indicate very high acid burden and a need for aggressive anti-reflux management.<sup>21</sup>

The optimal duration of reflux monitoring remains an important topic prompting debate among both investigators and clinicians. Although multiple studies discuss the augmented diagnostic yield of extending reflux monitoring to 96 hours<sup>22–25</sup>, lack of high quality evidence supporting the clinical value of this approach has not only limited clinical adoption, but has also been criticized as potentially misleading.<sup>4</sup> For the first time, this study provides the highest quality evidence that data from a 48-hour study (total AET 5.0% or 6.0%, or days of elevated AET) do not predict PPI discontinuation, because of unacceptably high rates of AET discordance between the first 24 hours of monitoring compared to the remainder of data. In contrast, the availability of additional days of recording beyond 48 hours mitigated the lack of reliability of day 1, and diagnostic performance actually improved with exclusion of day 1 from assessment of a 96-hour study. These findings demonstrate the discriminative value of a 96-hour study (over a 48-hour study) in predicting PPI discontinuation. Alternatively, a pragmatic approach with data assessment for at least 72 hours starting the morning following pH probe placement could be proposed, using total AET and abnormal days as preferred metrics.

The dual center, prospective, blinded design and use of successful PPI discontinuation as the study outcome are important strengths of this study. Overuse of PPIs is an important health care conundrum that requires objective documentation of lack of an acid-peptic mechanism for symptoms for successful reversal; this study establishes prolonged reflux



monitoring as the objective test of choice to make this determination. Although the best designed study to date for addressing important controversies, study limitations need to be considered. Each subject in our study served as their own comparator when comparing 48-hour to 96-hour monitoring, and while blinded randomization to 48-hour monitoring versus 96-hour monitoring could conceptually be viewed as more optimal, both subjects and study investigators were blinded to trial outcome and reflux monitoring results, serving to minimize potential risks of bias. Multiple measures were compared, which increases the potential of a type I error rate despite statistical use of mitigation methodology. While the study was conducted at tertiary care referral centers, the results should be generalizable to health care settings that manage patients with symptoms of GERD. Although not a direct limitation of the current study, GERD is a complex multifactorial disease with not just abnormal reflux burden, but also impaired mucosal defense mechanisms and altered gut-brain interactions.<sup>26</sup> Thus, acid burden alone can never completely address the pathogenic model of GERD or the mechanisms underlying symptom generation, leading to some patients with discordance between AET and PPI discontinuation, even at the extreme ends of the spectrum. Complementary esophageal physiologic tests evaluating mucosal integrity and baseline impedance could help better define the role of mucosal permeability and reflux hypersensitivity in the patients who were unable to stop PPI despite normal reflux burden.<sup>27-30</sup> We also could not demonstrate the value of reflux-symptom association to define a reflux sensitivity group, likely related to suboptimal patient documentation of symptoms, and to the fact that wireless monitoring does not capture weakly acidic or non-acidic reflux events, which may also generate symptoms. Finally, despite the high clinical relevance of PPI discontinuation as an outcome metric, the re-initiation of PPI therapy following reflux monitoring was a patient choice with multiple non-GERD influences including symptom burden, anxiety, hypervigilance, and fear of therapy. We hypothesize that an assessment of hypervigilance and visceral anxiety may help explain continued PPI use despite normal reflux burden, which could direct therapy toward behavioral interventions and neuromodulators.

In conclusion, we demonstrate that a 96-hour wireless study of acid suppression is more optimal than shorter durations of monitoring in predicting discontinuation vs. ongoing need for PPI therapy, using daily and composite AET of 4.0% as a physiologic threshold, and AET > 10.0% and/or DeMeester score > 50 as a marker for need for robust anti-reflux management measures. The most dispensable data are from day 1, and limiting data interpretation to the period starting the morning after probe placement could be a pragmatic interpretation approach that warrants further study. The hope is that these data will contribute to better clinical care of PPI non-responders and influence future research study design.

### Conflicts of Interest:

RY: Consultant for Medtronic, Phathom Pharmaceuticals, StatLinkMD; Research Support: Ironwood Pharmaceuticals; Advisory Board with Stock Options: RJS Mediagnostix

CPG: Consultant: Medtronic, Diversatek, Ironwood, Iso-Thrive, Quintiles

DAC: Consultant: Medtronic

PJK: Research support: Ironwood Pharmaceuticals; Advisory Board: Ironwood Pharmaceuticals, Johnson and Johnson, Reckitt

MFV: Consultant: Ironwood Pharmaceuticals, Diversatek, Phathom Pharmaceuticals; Daewood

Patent on mucosal integrity by Vanderbilt

JEP: Consultant: Medtronic, Ironwood Pharmaceuticals, Diversatek; Research support: Ironwood Pharmaceuticals, Takeda; Advisory Board: Medtronic, Diversatek; Stock Options: Crospon Inc

MM, BDN, JT, AJ, LK, AK: None

#### Research Funding Support:

This study was funded by NIH R01 DK092217-04 (PI: Pandolfino). RY is supported by NIH K23 DK125266 (PI: Yadlapati).

#### Abbreviations:

<b>PPI</b>	proton pump inhibitor
<b>GERD</b>	gastroesophageal reflux disease
<b>AET</b>	acid exposure time
<b>ROC</b>	receiver operating characteristics

#### REFERENCES

1. Gyawali CP, Carlson DA, Chen JW, et al. ACG Clinical Guidelines: Clinical Use of Esophageal Physiologic Testing. *Am J Gastroenterol* 2020;115:1412–1428. [PubMed: 32769426]
2. Zerbib F, Bredenoord AJ, Fass R, et al. ESNM/ANMS consensus paper: Diagnosis and management of refractory gastro-esophageal reflux disease. *Neurogastroenterol Motil* 2020:e14075. [PubMed: 33368919]
3. Gyawali CP, Kahrilas PJ, Savarino E, et al. Modern diagnosis of GERD: the Lyon Consensus. *Gut* 2018;67:1351–1362. [PubMed: 29437910]
4. Capovilla G, Salvador R, Spadotto L, et al. Long-term wireless pH monitoring of the distal esophagus: prolonging the test beyond 48 hours is unnecessary and may be misleading. *Dis Esophagus* 2017;30:1–8.
5. Roman S, Mion F, Zerbib F, et al. Wireless pH capsule--yield in clinical practice. *Endoscopy* 2012;44:270–6. [PubMed: 22275050]
6. Pandolfino JE, Kwiatek MA. Use and utility of the Bravo pH capsule. *J Clin Gastroenterol* 2008;42:571–8. [PubMed: 18364590]
7. Prakash C, Clouse RE. Value of extended recording time with wireless pH monitoring in evaluating gastroesophageal reflux disease. *Clin Gastroenterol Hepatol* 2005;3:329–34. [PubMed: 15822037]
8. Hirano I, Zhang Q, Pandolfino JE, et al. Four-day Bravo pH capsule monitoring with and without proton pump inhibitor therapy. *Clin Gastroenterol Hepatol* 2005;3:1083–8. [PubMed: 16271338]
9. Lacy BE, Dukowicz AC, Robertson DJ, et al. Clinical utility of the wireless pH capsule. *J Clin Gastroenterol* 2011;45:429–35. [PubMed: 21030875]
10. Abdallah J, George N, Yamasaki T, et al. Most Patients With Gastroesophageal Reflux Disease Who Failed Proton Pump Inhibitor Therapy Also Have Functional Esophageal Disorders. *Clin Gastroenterol Hepatol* 2019;17:1073–1080 e1. [PubMed: 29913281]
11. Spechler SJ, Hunter JG, Jones KM, et al. Randomized Trial of Medical versus Surgical Treatment for Refractory Heartburn. *N Engl J Med* 2019;381:1513–1523. [PubMed: 31618539]

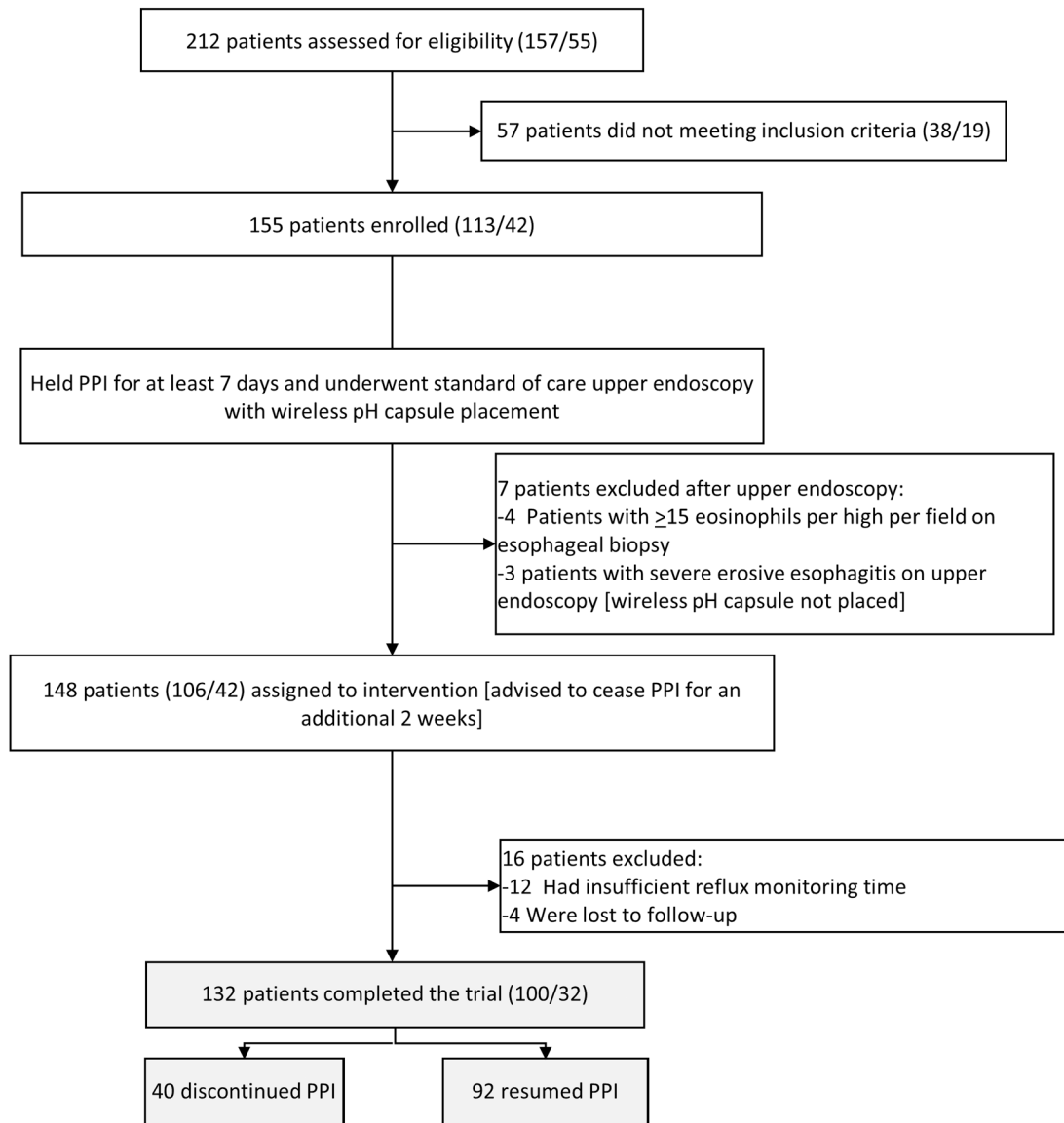
12. Yadlapati R, Masihi M, Gyawali CP, et al. Ambulatory Reflux Monitoring Guides Proton Pump Inhibitor Discontinuation in Patients With Gastroesophageal Reflux Symptoms: A Clinical Trial. *Gastroenterology* 2021;160:174–182 e1. [PubMed: 32949568]
13. Yadlapati R, Kahrilas PJ, Fox MR, et al. Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0((c)). *Neurogastroenterol Motil* 2021;33:e14058. [PubMed: 33373111]
14. Hasak S, Yadlapati R, Altayar O, et al. Prolonged Wireless pH Monitoring in Patients With Persistent Reflux Symptoms Despite Proton Pump Inhibitor Therapy. *Clin Gastroenterol Hepatol* 2020.
15. Yadlapati R, Ciolino JD, Craft J, et al. Trajectory assessment is useful when day-to-day esophageal acid exposure varies in prolonged wireless pH monitoring. *Dis Esophagus* 2019;32.
16. M S, E S. *epiR: Tools for the Analysis of Epidemiological Data*. R package version 2.0.26.. 2021.
17. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77. [PubMed: 21414208]
18. Pandolfino JE, Richter JE, Ours T, et al. Ambulatory esophageal pH monitoring using a wireless system. *Am J Gastroenterol* 2003;98:740–9. [PubMed: 12738450]
19. Wiener GJ, Morgan TM, Copper JB, et al. Ambulatory 24-hour esophageal pH monitoring. Reproducibility and variability of pH parameters. *Dig Dis Sci* 1988;33:1127–33. [PubMed: 3044715]
20. Ahlawat SK, Novak DJ, Williams DC, et al. Day-to-day variability in acid reflux patterns using the BRAVO pH monitoring system. *J Clin Gastroenterol* 2006;40:20–4. [PubMed: 16340628]
21. Krill JT, Naik RD, Higginbotham T, et al. Association Between Response to Acid-Suppression Therapy and Efficacy of Antireflux Surgery in Patients With Extraesophageal Reflux. *Clin Gastroenterol Hepatol* 2017;15:675–681. [PubMed: 27840185]
22. Scarpulla G, Camilleri S, Galante P, et al. The impact of prolonged pH measurements on the diagnosis of gastroesophageal reflux disease: 4-day wireless pH studies. *Am J Gastroenterol* 2007;102:2642–7. [PubMed: 17850412]
23. Penagini R, Sweis R, Mauro A, et al. Inconsistency in the Diagnosis of Functional Heartburn: Usefulness of Prolonged Wireless pH Monitoring in Patients With Proton Pump Inhibitor Refractory Gastroesophageal Reflux Disease. *J Neurogastroenterol Motil* 2015;21:265–72. [PubMed: 25843078]
24. Sweis R, Fox M, Anggiansah A, et al. Prolonged, wireless pH-studies have a high diagnostic yield in patients with reflux symptoms and negative 24-h catheter-based pH-studies. *Neurogastroenterol Motil* 2011;23:419–26. [PubMed: 21235685]
25. Connor J, Richter J. Increasing yield also increases false positives and best serves to exclude GERD. *Am J Gastroenterol* 2006;101:460–3. [PubMed: 16542280]
26. Niklasson A, Lindstrom L, Simren M, et al. Dyspeptic symptom development after discontinuation of a proton pump inhibitor: a double-blind placebo-controlled trial. *Am J Gastroenterol* 2010;105:1531–7. [PubMed: 20332770]
27. Gyawali CP, Tutuian R, Zerbib F, et al. Value of pH Impedance Monitoring While on Twice-Daily Proton Pump Inhibitor Therapy to Identify Need for Escalation of Reflux Management. *Gastroenterology* 2021;161:1412–1422. [PubMed: 34270955]
28. Patel DA, Higginbotham T, Slaughter JC, et al. Development and Validation of a Mucosal Impedance Contour Analysis System to Distinguish Esophageal Disorders. *Gastroenterology* 2019;156:1617–1626 e1. [PubMed: 30711626]
29. Rogers B, Samanta S, Ghobadi K, et al. Artificial intelligence automates and augments baseline impedance measurements from pH-impedance studies in gastroesophageal reflux disease. *J Gastroenterol* 2021;56:34–41. [PubMed: 33151406]
30. Zhang M, Pandolfino JE, Zhou X, et al. Assessing different diagnostic tests for gastroesophageal reflux disease: a systematic review and network meta-analysis. *Therap Adv Gastroenterol* 2019;12:1756284819890537.

**What is Known:**

- A significant portion of patients with esophageal reflux symptoms do not have GERD and can discontinue PPI therapy
- Wireless reflux monitoring off PPI is a standard diagnostic test for evaluation of GERD
- Diagnostic thresholds and optimal duration of wireless reflux monitoring are not well defined

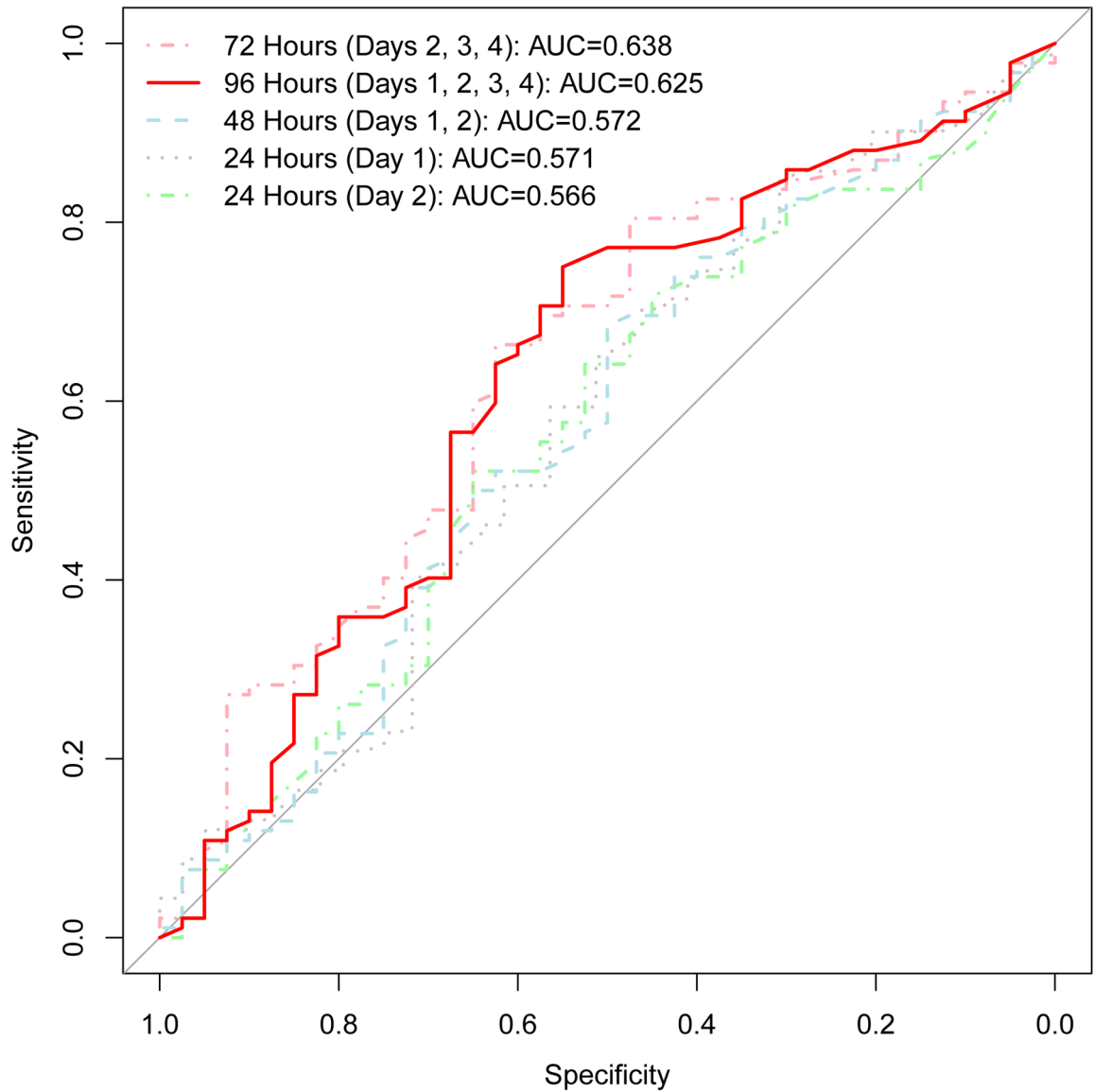
**What is New Here:**

- This double-blinded clinical trial included 132 adult patients with PPI non-responsive reflux symptoms undergoing prolonged wireless reflux monitoring
- Acid exposure time less than 4.0% was the optimal predictor of ability to discontinue PPI
- 96 hours of monitoring performed better than 48 hours in predicting ability to discontinue PPI



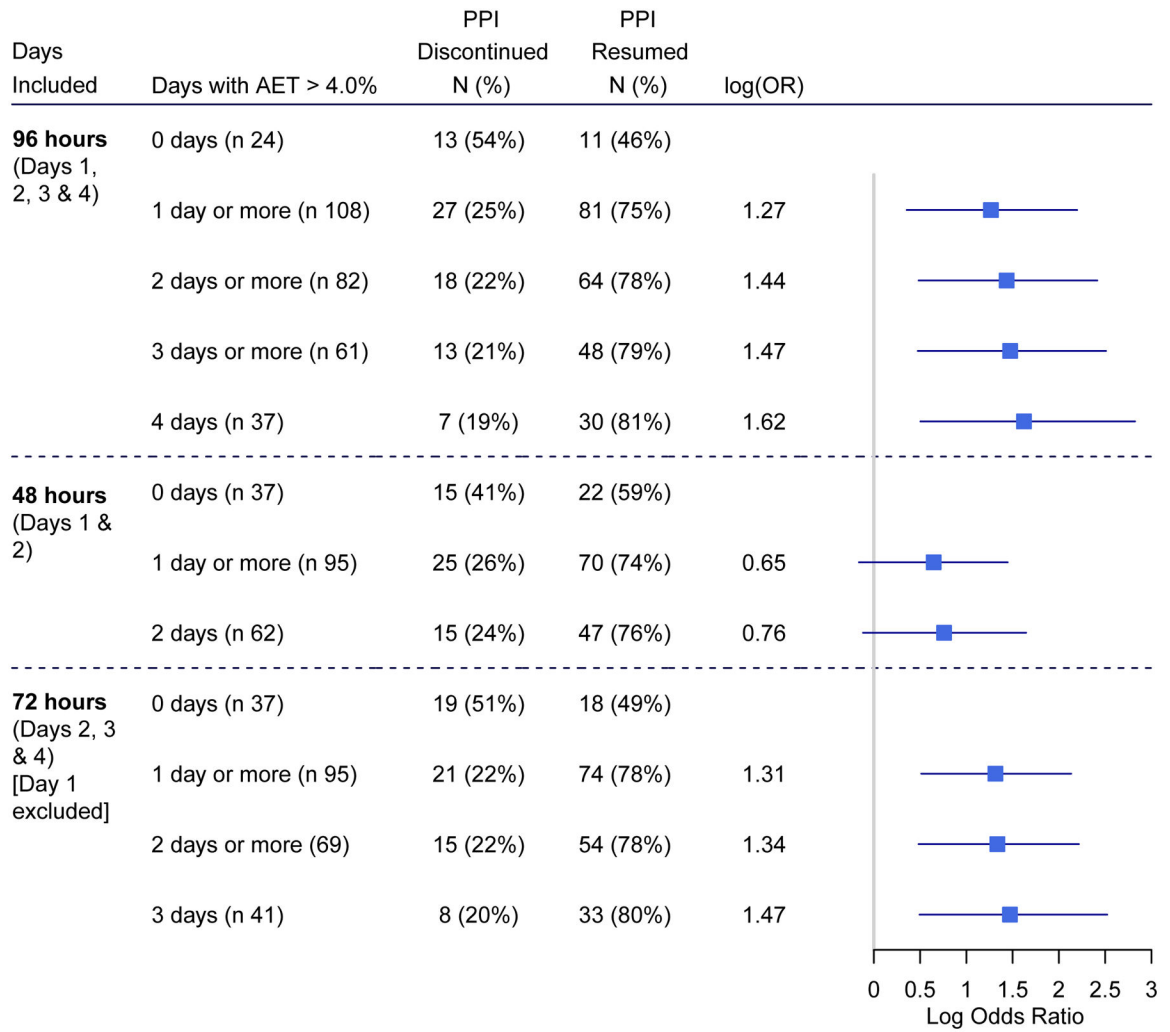
**Figure 1. Flow chart of study enrollment.**

Parenthesis indicate site (patients from Northwestern University/patients from Washington University).



**Figure 2. Receiver operating characteristics for an acid exposure time threshold of 4.0% in predicting ability to discontinue PPI for four different durations of reflux monitoring off PPI.** Area under the curve (AUC) for 96 hours of monitoring is significantly greater than that of 24 hours ( $p=0.03$ ) and that of 48 hours ( $p=0.10$ ). AUC is similar for 24 hours (day 1) and 48 hours (days 1 and 2) of monitoring ( $p=0.49$ ). AUC is similar for 96 hours total and 96 hours excluding day 1 (days 2, 3, and 4) of monitoring ( $p=0.43$ ).





**Figure 3.** Forest plot comparing risk estimates for predicting ability to discontinue PPI between 96-hour, 48-hour, and 96-hour with day 1 excluded monitoring

**Table 1.**

Risk estimates for data from 96-hour wireless pH monitoring off PPI in predicting ability to discontinue PPI

Metrics	Category	PPI Discontinued	PPI Resumed	OR (95% CI)	p-value
<b>Total AET 4.0%</b>	Total AET 4.0% (n 49)	22 (45%)	27 (55%)	Reference	
	Total AET > 4.0% (n 83)	18 (22%)	65 (78%)	0.34 (0.16, 0.73)	0.006
<b>Total AET 5.0%</b>	Total AET 5.0% (n 62)	25 (40%)	37 (60%)	Reference	
	Total AET > 5.0% (n 70)	15 (21%)	55 (79%)	0.40 (0.18, 0.86)	0.020
<b>Total AET 6.0%</b>	Total AET 6.0% (n 72)	27 (38%)	45 (63%)	Reference	
	Total AET > 6.0% (n 60)	13 (22%)	47 (78%)	0.46 (0.21, 0.99)	0.051
<b>* Days with AET &gt; 4.0%</b>	0 days (n 24)	13 (54%)	11 (46%)	Reference	
	1 day or more (n 108)	27 (25%)	81 (75%)	0.28 (0.11, 0.70)	0.007
	2 days or more (n 82)	18 (22%)	64 (78%)	0.24 (0.09, 0.62)	0.003
	3 days or more (n 61)	13 (21%)	48 (79%)	0.23 (0.08, 0.62)	0.004
	4 days (n 37)	7 (19%)	30 (81%)	0.20 (0.06, 0.60)	0.006
<b>* Days with AET &gt; 5.0%</b>	0 days (n 31)	16 (52%)	15 (48%)	Reference	
	1 day or more (n 101)	24 (24%)	77 (76%)	0.29 (0.12, 0.68)	0.004
	2 days or more (n 72)	16 (22%)	56 (78%)	0.27 (0.11, 0.65)	0.004
	3 days or more (n 52)	12 (23%)	40 (77%)	0.28 (0.11, 0.72)	0.009
	4 days (n 27)	4 (15%)	23 (85%)	0.16 (0.04, 0.54)	0.005
<b>* Days with AET &gt; 6.0%</b>	0 days (n 38)	18 (47%)	20 (53%)	Reference	
	1 day or more (n 94)	22 (23%)	72 (77%)	0.34 (0.15, 0.75)	0.008
	2 days or more (n 62)	14 (23%)	48 (77%)	0.32 (0.13, 0.77)	0.011
	3 days or more (n 39)	10 (26%)	29 (74%)	0.38 (0.14, 0.98)	0.050
	4 days (n 19)	4 (21%)	15 (79%)	0.30 (0.07, 0.99)	0.061
<b>DeMeester Score 14.2</b>	DeMeester Score 14.2 (n 28)	13 (46%)	15 (54%)	Reference	
	DeMeester Score > 14.2 (n 89)	22 (25%)	67 (75%)	0.38 (0.16, 0.92)	0.032
<b>DeMeester Score 50</b>	DeMeester Score 50 (n 97)	33 (34%)	64 (66%)	Reference	
	DeMeester Score > 50 (n 20)	2 (10%)	18 (90%)	0.22 (0.03, 0.81)	0.048
<b>* Dominant Pattern</b>	Physiologic (n 65)	26 (40%)	39 (60%)	Reference	
	Pathologic (n 58)	13 (22%)	45 (78%)	0.43 (0.19, 0.94)	0.038
	Borderline/Discordant (n 9)	1 (11%)	8 (89%)	0.19 (0.01, 1.11)	0.125
<b>* Trajectory Pattern</b>	Low (n 71)	27 (38%)	44 (62%)	Reference	
	Middle (n 52)	11 (21%)	41 (79%)	0.44 (0.19, 0.97)	0.048
	High (n 9)	2 (22%)	7 (78%)	0.47 (0.07, 2.10)	0.362
<b>Symptom Index</b>	Symptom Index < 50 (n 92)	30 (33%)	62 (67%)	Reference	
	Symptom Index 50 (n 40)	10 (25%)	30 (75%)	0.69 (0.29, 1.56)	0.383
<b>Symptom Association Probability</b>	SAP < 95 (n 75)	27 (36%)	48 (64%)	Reference	
	SAP 95 (n 57)	13 (23%)	44 (77%)	0.53 (0.24, 1.13)	0.105

\* Note that risk estimates are for PPI discontinuation with reference to the lowest acid burden group. For these an OR less than 1.0 with a 95% CI not crossing 1.0 indicates a significantly lower odds of PPI discontinuation.

**Table 2.**

Sensitivity/Specificity for PPI Resumption for AET Thresholds and Days Positive

		Acid Exposure Time Threshold		
		AET 4.0%	AET 5.0%	AET 6.0%
Number of days above AET threshold	0 Days	0.120 / 0.675	0.163 / 0.600	0.217 / 0.550
	1+ Days	0.880 / 0.325	0.837 / 0.400	0.783 / 0.450
	2+ Days	0.696 / 0.550	0.609 / 0.600	0.522 / 0.650
	3+ Days	0.522 / 0.675	0.435 / 0.700	0.315 / 0.750
	4+ Days	0.326 / 0.825	0.250 / 0.900	0.163 / 0.900

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**

Risk estimates for data from 48-hour wireless pH monitoring off PPI in predicting ability to discontinue PPI

Metrics	Category	PPI Discontinued	PPI Resumed	OR (95% CI)	p-value
<b>Total AET 4.0%</b>	Total AET 4.0% (n 49)	20 (41%)	29 (59%)	Reference	
	Total AET > 4.0% (n 83)	20 (24%)	63 (76%)	0.45 (0.21, 1.00)	0.045
<b>Total AET 5.0%</b>	Total AET 5.0% (n 61)	21 (34%)	40 (66%)	Reference	
	Total AET > 5.0% (n 71)	19 (27%)	52 (73%)	0.71 (0.33, 1.43)	0.340
<b>Total AET 6.0%</b>	Total AET 6.0% (n 74)	26 (35%)	48 (65%)	Reference	
	Total AET > 6.0% (n 58)	14 (24%)	44 (76%)	0.59 (0.27, 1.25)	0.174
<b>* Days with AET &gt; 4.0%</b>	0 days (n 37)	15 (41%)	22 (60%)	Reference	
	1 day or more (n 95)	25 (26%)	70 (74%)	0.52 (0.24, 1.17)	0.113
	2 days (n 62)	15 (24%)	47 (76%)	0.47 (0.19, 1.12)	0.090
<b>* Days with AET &gt; 5.0%</b>	0 days (n 45)	18 (40%)	27 (60%)	Reference	
	1 day or more (n 87)	22 (25%)	65 (75%)	0.51 (0.23, 1.10)	0.084
	2 days (n 49)	12 (25%)	37 (76%)	0.49 (0.20, 1.17)	0.110
<b>* Days with AET &gt; 6.0%</b>	0 days (n 54)	21 (39%)	33 (61%)	Reference	
	1 day or more (n 78)	19 (24%)	59 (76%)	0.51 (0.24, 1.07)	0.076
	2 days (n37)	10 (27%)	27 (73%)	0.58 (0.23, 1.42)	0.243

\* Note that risk estimates are for PPI discontinuation with reference to the lowest acid burden group. For these an OR less than 1.0 with a 95% CI not crossing 1.0 indicates a significantly lower odds of PPI discontinuation.

**Table 4.**

Risk estimates for data from 96-hour (minus day 1) wireless pH monitoring off PPI in predicting ability to discontinue PPI

Metrics	Category	PPI Discontinued	PPI Resumed	OR (95% CI)	p-value
<b>Total AET 4.0%</b>	Total AET 4.0% (n 55)	24 (44%)	31 (56%)	Reference	
	Total AET > 4.0% (n 77)	16 (21%)	61 (79%)	0.33 (0.15, 0.71)	0.006
<b>Total AET 5.0%</b>	Total AET 5.0% (n 65)	26 (40%)	39 (60%)	Reference	
	Total AET > 5.0% (n 67)	14 (21%)	53 (79%)	0.40 (0.18, 0.83)	0.018
<b>Total AET 6.0%</b>	Total AET 6.0% (n 81)	29 (36%)	52 (64%)	Reference	
	Total AET > 6.0% (n 51)	11 (22%)	40 (78%)	0.50 (0.21, 1.11)	0.086
<b>* Days with AET &gt; 4.0%</b>	0 days (n 37)	19 (51%)	18 (49%)	Reference	
	1 day or more (n 95)	21 (22%)	74 (78%)	0.27 (0.12, 0.60)	0.001
	2 days or more (69)	15 (22%)	54 (78%)	0.26 (0.11, 0.62)	0.002
	3 days (n 41)	8 (20%)	33 (81%)	0.23 (0.08, 0.61)	0.004
<b>* Days with AET &gt; 5.0%</b>	0 days (n 46)	22 (48%)	24 (52%)	Reference	
	1 day or more (n 86)	18 (21%)	68 (79%)	0.29 (0.13, 0.62)	0.002
	2 days or more (n 61)	14 (23%)	47 (77%)	0.32 (0.14, 0.74)	0.008
	3 days (n 32)	5 (16%)	27 (84%)	0.20 (0.06, 0.58)	0.005
<b>* Days with AET &gt; 6.0%</b>	0 days (n 55)	24 (44%)	31 (56%)	Reference	
	1 day or more (n 77)	16 (21%)	61 (79%)	0.34 (0.16, 0.72)	0.006
	2 days or more (n 48)	13 (27%)	35 (73%)	0.48 (0.20, 1.09)	0.083
	3 days (n 27)	5 (19%)	22 (82%)	0.29 (0.09, 0.84)	0.030

\* Note that risk estimates are for PPI discontinuation with reference to the lowest acid burden group. For these an OR less than 1.0 with a 95% CI not crossing 1.0 indicates a significantly lower odds of PPI discontinuation.